

Developmental shift in multimodal cue usage in early mother-child conversational initiation

Jun Ho Chai¹, Barbara Zapiór², Eon-Suk Ko¹

¹ Center for Data Science in Humanities, Chosun University, Gwangju, Korea;

² School of Psychology, University of East Anglia, Norwich, UK

Abstract

Early mother-child conversational initiation is increasingly understood as a developmentally dynamic, embodied, and socially embedded process in which caregivers and children coordinate multimodal cues, including gaze, gesture, touch, and speech, to establish joint attention and meaning. Drawing on an existing audio-visual corpus of Korean mother-child interactions from 35 dyads aged 8 to 27 months, sampled across stages of language and motor development, we examine how nonverbal cues preceding verbal conversational onset vary with child age, sex, and speaker role. Results show that initiatory multimodal cues decline with age, especially in conversational blocks where mothers initiate the vocal exchanges, a pattern consistent with a developmental shift toward more verbally-driven interactional strategies. Look cues and object-oriented tactile cues were the most frequent signals preceding conversational blocks, highlighting their central role in initiating exchanges. In addition, initiators produce more initiatory multimodal cues than their partners in the moments preceding conversational onset, pointing to a strategic use of multimodal signals to capture attention and frame the upcoming verbal interaction. Overall, the findings highlight the key role of gaze and object-oriented action in early interaction and the

adaptability of communicative strategies in mother–child exchanges as children take on a greater role in verbal initiation.

Keywords: Multimodal Communication, Nonverbal Cues, Mother-Child Interaction, Conversational Exchanges, Developmental Change

Introduction

The *initiation* of conversational exchanges between caregivers and children is a cornerstone of early social interaction and language development. While verbal communication is often regarded as the primary channel for initiating conversation in early childhood (Chai et al., 2025; Ko et al., 2016; Salo et al., 2022), infants and caregivers frequently rely on nonverbal signals such as gestures (Caselli et al., 2012; Kuhn et al., 2014; Shin et al., 2024), facial expressions (Chong et al., 2003; Schachner, 2022), and touch (Abu-Zhaya et al., 2017; Tincoff et al., 2019; Ko et al., 2023) long before fluent speech emerges. These multimodal cues capture attention, establish joint focus, and regulate turn-taking, making them a critical part of early communicative exchanges. Despite broad agreement that multimodal signals play a non-trivial role in early communication, it is not well understood how particular nonverbal cues are used to initiate interaction, how their use changes over development, or how the speaker-respondent role in conversational interactions shapes these processes. The present study addresses this gap by providing a fine-grained developmental analysis of how mothers and children use gaze, gesture, and touch to initiate conversational exchanges in naturalistic interaction.

Multimodal foundations of early conversation initiation

Caregivers employ a wide range of multimodal cues, including gaze, gestures, and visual, auditory, and tactile signals, to engage with their children and support their

communicative development (Tomasello, 1988; Mundy & Newell, 2007; Abu-Zhaya et al., 2017; Ko et al., 2023; Koise & Lew-Williams, 2024). The use of such multimodal communication strategies likely contributes to the effectiveness of adult-initiated interactions in promoting language development. Studies have investigated how caregivers combine verbal input with visual and gestural cues (Bruner, 1983), maintain temporal synchrony in verbal and gestural communication during interactions (Gogate et al., 2000; Vigliocco et al., 2014), and adapt their use of visual, auditory, and tactile cues to the child's development (Ko et al., 2023). Eye gaze, in particular, supports joint attention and referential clarity, while physical touch often plays a critical role in regulating attention, affect, and turn-taking (Çetinçelik et al., 2021; Csibra, 2010; Gredebäck et al., 2010; Yu & Ballard, 2007; Kendon, 1967).

Children, too, contribute actively to early exchanges by utilizing a growing repertoire of multimodal cues. From early infancy, children develop an action-gesture "vocabulary" to signal attention and intention to caregivers (Caselli et al., 2012). As motor and cognitive capacities develop, new behaviors such as pointing (Behne et al., 2012; Moore & D'Entremont, 2001) and coordinated eye contact (Yurkovic et al., 2021) emerge, enabling richer interactional participation. These abilities evolve through developmental milestones. For instance, walking has been linked to shifts in communication style and engagement patterns (West & Iverson, 2021). Despite this extensive use of multimodal signals, little is known about how specific cues, such as gestures versus touch, are differentially employed to initiate interaction and whether certain cue types are more prevalent at different stages of development.

Developmental change and initiator roles in multimodal cue usage

The nature and frequency of specific cue types, such as gaze, gesture, or touch, are not static but evolve with child development. Ko et al. (2023) found that mothers adjust their

use of touch across developmental stages as infants become more mobile and vocal. Similarly, Gogate et al. (2000) demonstrated that caregivers synchronize multimodal cues, such as speech and gesture, more precisely as infants' attention and learning capacities develop. These findings suggest that multimodal cue usage is developmentally sensitive, and tracking how such cues change over time can offer crucial insights into how children transition from sensorimotor to symbolic modes of interaction. What remains underspecified, however, is how these developmental changes are reflected specifically in the moment of conversational initiation.

In addition to cue usage, who initiates the interaction also shapes the structure and dynamics of early exchanges. While adult-initiated interactions are often positively associated with language outcomes (Salo et al., 2022), children also appear to be more engaged when they take the lead (Chai et al., 2025; Ko et al., 2016). Research has identified an "initiator effect" in both English- (greater segment duration and shorter response times in self-initiated conversation blocks; Ko et al., 2016) and Korean-speaking dyads (greater segment number and duration, greater turn number, and shorter response times in self-initiated conversation blocks; Chai et al., 2025). These findings indicate heightened engagement in self-initiated conversations by both adults and children in a naturalistic environment. Importantly, initiator identity may influence not only the verbal structure of the interaction but also the accompanying nonverbal behaviors. Yet little is currently known about how initiators (i.e., mothers and children) use multimodal cues, such as gaze, gesture, or touch, during interaction initiation, and whether these cues differ depending on who initiates.

Theoretical perspectives on embodied and distributed interaction

Early mother–child multimodal communication can be understood within several theoretical perspectives explaining how interaction supports language development and

cognition. Classic social-interactionist accounts of language acquisition (Bruner, 1983; Vygotsky & Cole, 1978) emphasize the role of social interaction, caregiver scaffolding, and multimodal cues in supporting children's emerging linguistic capacities. These perspectives highlight the multimodality of early linguistic environments (Custode & Tamis-LeMonda, 2020; Gogate et al., 2000; Tamis-LeMonda et al., 2013) and the importance of joint attention for establishing shared referential focus (Tomasello, 1988). Seminal research on gaze direction (Kendon, 1967) further demonstrated how nonverbal cues signal engagement and coordinate interaction between partners, a focus that remains significant today when considering how conversational exchanges are initiated and by whom.

Building on these interactionist insights, contemporary approaches view cognition as distributed across bodies, tools, and environments (Newen et al., 2018; Tamis-LeMonda & Masek, 2023; Vesper & Sevdalis, 2020). The 5E cognition framework, encompassing embodied, embedded, enacted, extended, and ecological approaches, provides a foundation for understanding how conversational initiation emerges through dynamic coupling between brain, body, and environment rather than isolated neural processes (Gallagher, 2023). In this view, conversational initiation is not reducible to internal cognitive processes but arises from dynamic sensorimotor coordination embedded within real-time, material contexts.

Communicative exchanges are embodied (shaped by bodily capacities for gaze, gesture, and touch), embedded (situated within rich social and material contexts), enacted (emerging through active engagement rather than passive processing), extended (incorporating objects as cognitive resources), and ecologically situated (constrained and afforded by environmental structures; Newen et al., 2018; Rietveld et al., 2018). Material Engagement Theory (Malafouris, 2013, 2020) further argues that interactions with objects are not merely supportive of cognition but constitutive of it. Joint attention episodes involving shared gaze and object manipulation are not just communication strategies but expressions of cognition-

in-action. As children explore and act on their environments, they draw meaning from material affordances in ways that shape their communicative and cognitive development.

A rich tradition in sociocultural, gesture-focused, and multimodal communication research has long demonstrated that conversational initiation is fundamentally embodied, culturally shaped, and mediated through repeated interactional routines. Language socialization research (Ochs, 1982, 1992; Schieffelin, 1990; Snow, 1977) reveals how caregivers scaffold children into culturally specific ways of entering interaction through participation formats and guided routines, establishing norms about who initiates and how. Simultaneously, conversation-analytic and gesture research (Goodwin, 2000, 2007; Kendon, 2004) has shown that speakers coordinate gaze, gesture, body orientation, and objects to secure reciprocity and launch conversational turns framed as multimodal interaction during infancy. Work by Goldin-Meadow and colleagues (Iverson & Goldin-Meadow, 2005; Goldin-Meadow et al., 2007, 2009) further established that early gestures, particularly deictic gestures (pointing and showing), are developmental precursors to speech and revealing indicators of children's readiness to communicate, motivating attention to gestural and referential behaviors in early conversational initiation. Recent European scholarship on early communicative development (Murillo & Casla, 2013; Moreno-Núñez et al., 2015; Rodríguez & Moreno-Núñez, 2012) has documented how caregivers and infants organize joint attention and participation through object-mediated communicative routines and ostensive gestures – a perspective directly aligned with the prevalence of object-related tactile cues and gaze in interaction initiation. Building on these empirically grounded traditions, contemporary accounts of embodied and distributed cognition (Newen et al., 2018; Gallagher, 2023; Malafouris, 2013, 2020), including 5E cognition and Material Engagement Theory, provide a conceptual vocabulary for interpreting how conversational initiation emerges through dynamic coordination of bodies, objects, and cultural contexts. Importantly, while these

frameworks help organize and interpret patterns of multimodal behavior, the present study contributes behavioral and developmental evidence of how such embodied, culturally situated practices unfold across early childhood, rather than direct evidence of extended or constitutive cognitive processes.

The present study

Despite recognition of these theoretical perspectives, not much is known about how multimodal cues are used specifically to initiate conversations, how their use changes over development, or how initiator roles shape multimodal strategies. Few studies have examined these dynamics using naturalistic data within a framework that integrates embodiment, development, and material engagement (Tomasello et al., 2005). The current study addresses these gaps by pursuing three research questions: (1) Which multimodal cues are most frequently used by mother–child dyads to initiate conversational exchanges? (2) How does the usage of these multimodal cues shift across the course of development, particularly as children transition from preverbal to early verbal stages? (3) Does the identity of the initiator influence the frequency and type of cues, with different patterns emerging when the mother versus the child initiates the interaction?

To address these questions, the present study adopts a fine-grained behavioural coding approach to examine how multimodal cues are deployed in the moments leading up to conversational initiation. The *initiator* of a conversational block is defined as the participant (mother or child) who produces the first utterance of that block, based purely on the vocalization. Nonverbal behaviors were coded within the 3-second window preceding this first utterance but did not determine initiator status, allowing examination of multimodal signals deployed before vocal initiation. As little is currently known about the optimal length

of a coding window that best captures multimodal behaviors preceding speech onset, we utilized a 3-second window. This duration is commonly considered sufficient for examining parents' responses that follow infants' signals in naturalistic play interactions (Bornstein et al., 2019; Moreno-Núñez et al., 2021; Van Egeren et al., 2001), and thus provides a practical and comparable basis for the analyses in the present study. To answer the research questions, we analyzed an existing corpus of Korean mother–child interaction (Ko et al., 2020) comprising cross-sectional data of 35 mother-child dyads with children aged between 8 and 27 months, using a fine-grained behavioral coding approach to annotate cue usage patterns across age, initiator role, and cue type. While the design provides developmental snapshots that enable us to investigate age-related patterns, it does not trace individual developmental trajectories over time.

Methods

Data

Data used for analysis originated from an existing corpus of Korean mother–child interaction (reference redacted for review). The corpus consists of audio-visual recordings of 35 naturalistic mother–child interactions collected in a mock apartment designed to resemble a typical household living space for families with young children (Figure 1). The dyads included mothers with their children aged 8 months ($M = 7.9$, $SD = 1.3$, $Min = 6$, $Max = 9.7$; $N = 12$, 4 girls), 14 months ($M = 13.8$, $SD = 1.3$, $Min = 12$, $Max = 16$; $N = 11$, 6 girls), and 27 months ($M = 26.7$, $SD = 1.6$, $Min = 25$, $Max = 30$; $N = 12$, 4 girls). The mothers were generally highly educated: 2 had completed high school, 23 held college degrees, and 11 had graduate-level education, which is above the national average of college attendance (69.7%). Their ages ranged from 28 to 41, with a mean age of 33. The apartment included a bedroom, a kitchen, a foyer, and a living room with a play area where the dyads spent the majority of the session. The mothers were instructed to interact with their children naturally, similar to

their typical home routines. All study procedures involving human participants were approved by the Institutional Review Board of [redacted for review] (Approval No. 1602/001-007).

Figure 1. Layout of the mock apartment used for recording the corpus data. Adapted from a previously published dataset (details redacted for review). CC BY license.



Although this study used a fixed, pre-existing dataset, it remains important to evaluate whether the sample size was adequate for detecting developmental patterns. Because no prior studies have directly examined multimodal cues preceding conversational initiation, we drew on a related analysis of touch–speech co-occurrence in caregiver input (Author, 2023) as an approximate benchmark. That study found clear developmental differences in how often touch co-occurred with speech across age groups. Based on the reported group means and standard deviations, we estimated an approximate effect size, suggesting a large effect ($\eta^2 \approx .39$, $f \approx 0.80$). Based on this effect size, a one-way ANOVA with three groups, $\alpha = .05$, and power = .80 would require 18 participants (6 per group; G*Power 3.1; Faul et al., 2007). Our dataset of 35 dyads therefore exceeds this threshold, providing sufficient power to detect medium-to-large

effects, although actual effect sizes for cue initiation may differ because the present study examines a broader set of multimodal behaviors.

The recordings were captured by multiple cameras in the apartment, while the audio was recorded using clip-on microphones on vests worn by both the mother and child. While the video recordings consisted of approximately 40-minute naturalistic mother-child play interactions and a 10-minute conversation with an adult, the segment interacting with an adult was not analyzed due to our specific interest in dyadic interactions. One additional recording is part of the corpus, but the data was not analyzed due to the unavailability of a speech transcript because of the child's constant crying.

Coding nonverbal cues

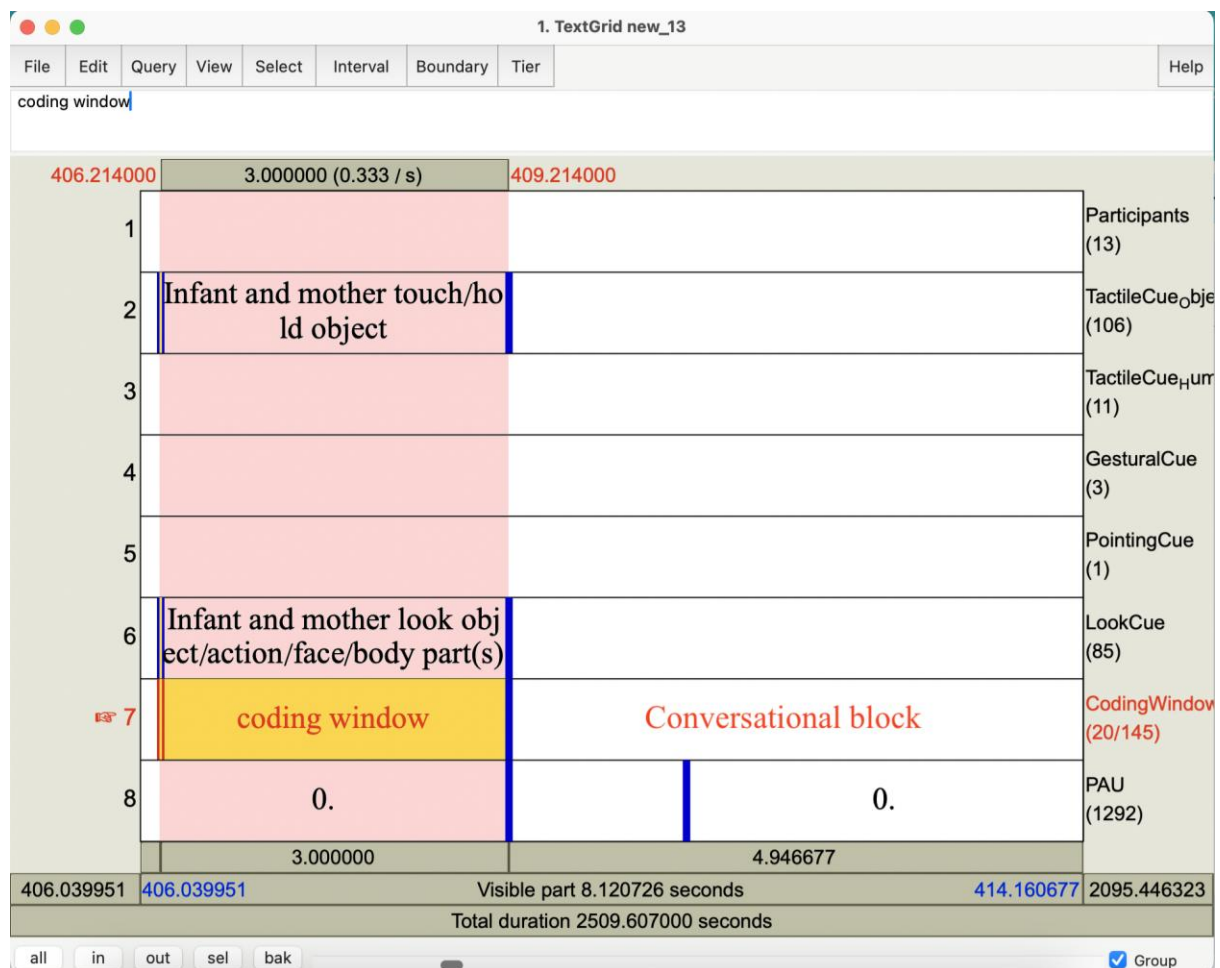
All maternal and child vocalizations were drawn from the existing transcriptions and utterance segmentations of the Ko corpus, which follow established CHILDES conventions (Ko et al., 2020). Within this framework, child productions that were judged to be communicative (including early word forms and proto-words) are represented as individual utterances. We used a custom R script to derive conversational blocks and coding windows based on the existing time-aligned transcripts of the Korean mother-child corpus. Following the LENA convention, the script grouped vocalizations from any speaker into conversational blocks separated by pauses of 5 seconds or longer using the pause tier in the transcripts to identify silence intervals. For each block, the script then marked a 3-second window immediately preceding the onset of the first utterance (the *initiator*) to capture potential nonverbal cues, and exported the resulting block boundaries and windows in Praat TextGrid format. These TextGrid files were then imported into ELAN together with the corresponding video recordings, which contained synchronized audio for manual annotation of multimodal cues.

For each conversational block, the term initiator was defined as the participant who produced the first utterance of that block (mother or child). Nonverbal cues, including mutual tactile cues (e.g., the infant and mother touching each other or the same object), were coded within the 3-second window preceding the onset of the first utterance, but they did not determine initiator status. Mutual touch and other joint cues could occur before either mother-initiated or child-initiated blocks and are interpreted as features of the pre-vocal interactional context rather than as establishing who initiated the conversation.

Within ELAN, we created standardized templates for coding non-verbal cues by adding multiple tiers, each corresponding to a specific cue category (see Table 1). These tiers included participant visibility (whether within or outside the video frame), tactile cues (human or object), gestures, pointing, and gaze (look) behaviors. Coders annotated behaviors directly on these tiers using predefined drop-down menus, which reduced labeling errors and ensured consistency across sessions. Using this approach, we identified and coded non-verbal cues that occurred within the 3-second window preceding each conversational block (Figure 2).

Figure 2.

Example of a conversational block and nonverbal cues identified based on predefined ethograms. Note that the PAU tier represents silence, in which 0 represents non-speech pauses between utterances.



Two trained coders annotated the non-verbal cues within each coding window. Both coders underwent initial training on the coding scheme, including review of the cue definitions and practice sessions using sample recordings not included in the analysis dataset. To assess inter-rater reliability, we selected a subset of recordings that were independently annotated by both coders. Reliability checks were performed in two passes to allow for iterative feedback and refinement of coding practices. In the first pass, agreement between coders reached Cohen's $\kappa = 0.774$ ($z = 23.8$, $p < .001$), with 91.3% raw agreement, indicating “substantial” reliability according to Landis and Koch's (1977) benchmark. However, one coder produced a relatively high number of annotations, particularly for ambiguous or borderline movements, which inflated coding density. After reviewing these discrepancies, both coders agreed to emphasize the intended communicative function of a movement or

touch rather than any incidental motion (e.g., shifting body position without clear communicative intent). Following this calibration, a second pass of reliability coding was conducted on a separate subset of recordings. Agreement improved to Cohen's $\kappa = 0.806$ ($z = 24.8, p < .001$), with 92.7% raw agreement, reflecting "almost perfect" reliability. After establishing this level of consistency, the primary coder proceeded to annotate all remaining recordings.

Coding criteria¹ (Table 1) were established for various cue types, including look cues, tactile cues (object and human), gestural cues, and pointing cues. Look cues were coded when the infant and/or mother looked at an object, action, face, or body part, with the initiator of the conversational block being noted. Tactile cues were divided into object and human categories, with object tactile cues involving intentional touching or holding of objects by the infant and/or mother, and human tactile cues involving intentional touching between the infant and mother. Gestural cues were coded when the infant and/or mother produced visible movements directed toward an object or action within proximal reach, excluding instances of touching or holding, which were coded as tactile cues, as well as pointing gestures, which were coded separately due to their distinct attentional properties². Additionally, we coded for participant visibility, documenting instances where the mother, infant, or both were outside the video frame.

Table 1.

¹ Nonverbal behaviors including mutual tactile cues (e.g., the infant and mother touching each other or the same object), were coded within the 3-second window preceding the onset of the first utterance but did not determine initiator status. As a result, mutual touch and other joint cues could occur before either mother-initiated or child-initiated blocks and are interpreted as features of the pre-vocal interactional context rather than as establishing who initiated the conversation. This distinction allows us to examine whether the person who is about to speak deploys multimodal cues strategically in the moments before vocal initiation.

² Note that ostensive gestures such as showing and giving were distributed across the gestural and tactile-object categories: the pre-contact phase (lifting an object toward the partner, offering motion) was coded as a gestural cue, while sustained holding, transfer, or contact with the partner transitioned the behavior to tactile-object engagement. This distribution reflects our coding priority to distinguish the attention-directing function of gesture from the material-engagement function of tactile contact. However, we acknowledge that ostensive actions span both categories and may be underrepresented as a unified phenomenon in this analysis, and accordingly we have dedicated a paragraph in the Discussion to their developmental significance in light of Guevara et al. (2024) and related work.

Coded Cue Categories and Subcategories with Acronyms

Cue Category	Cue Subcategory
LookCue	Mother look object/action/face/body part(s) (initiator of CB)
	Mother following infant look object/action/face/body part(s)
	Infant look object/action/face/body part(s) (initiator of CB)
	Infant following mother look object/action/face/body part(s)
	Infant and mother look object/action/face/body part(s)
TactileCue_Object	Mother touch/hold object
	Infant touch/hold object
	Infant and mother touch object
TactileCue_Human	Mother touch infant
	Infant touch mother
	Infant and mother touch each other
GesturalCue	Mother gesture to object/action
	Infant gesture to object/action
	Infant and mother gesture object/action
PointingCue	Mother point to object/action
	Infant point to object/action
	Infant and mother point to object/action

Coders adjusted or skipped coding windows when necessary. This flexibility was primarily required in two situations: when silence or vocalization periods did not align precisely with the predefined coding window, and when non-speech sounds (TV sound or phone call) initiated a conversational block without subsequent vocalizations. This involved checking the corresponding PAU tier for silences between 4 and 5 seconds and adjusting the boundaries accordingly. Coding windows were skipped if the adjusted silence duration was less than 5 seconds.

For each segment of the conversation, the dataset logs the identity of the participant, the segment's start and end times, and specifies their engagement during that period. Nonverbal communicative behaviors are systematically annotated in detail, capturing presence and timing of specific categories such as manual gestures (e.g., hand or arm movements), body movements, gaze direction (including mutual or averted gaze), pointing actions, and any contact with people or objects. Each nonverbal behavior is categorized according to a predefined coding scheme developed for the study. In parallel, the dataset documented every instance of vocalization, recording the precise onset and offset of each spoken utterance, and indicating whether the speaker was the child or the mother. This annotation enables alignment between verbal and nonverbal behaviors at the level of individual moments within the interaction. For any segment in which no target nonverbal behaviors are observed, this is explicitly coded and the resulting dataset provides a continuous, segment-by-segment record of both verbal and nonverbal communicative behavior.

Data Analysis

To address the research questions with the presence of cues as the outcome variable, we constructed a series of binomial linear mixed models, using the *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), *emmeans* (Lenth, 2021), *ggeffects* (Lüdtke, 2018), and *modelbased* (Makowski et al., 2025) packages in *R* (R Core Team, 2021).

To characterize nonverbal cue patterns developmentally, we first coded cues using fine-grained multimodal categories, allowing us to detect category-specific patterns across age groups that may align with distinct communicative functions or developmental milestones. Age was initially modeled as a categorical variable (months), capturing broad group effects that reflect major developmental stages and provide summary-level contrasts between groups. In subsequent analyses, to more closely examine the patterns of

developmental change, age was treated as a continuous variable (months), permitting finer-grained modeling of within-year shifts and nonlinear changes across the sampled age range. To maximize statistical power and interpretability in these more nuanced models, we then collapsed cue types to focus on overall cue presence or absence.

Following established practices in developmental research, we included child sex as a covariate in all models to account for known sources of variance in early language outcomes. We employed a systematic model comparison approach using likelihood ratio tests and information criteria (AIC, BIC) to identify the most parsimonious model. We also conducted a post-hoc sensitivity analysis to characterise the smallest fixed-effect patterns that could be detected with the available data. Approximating the mixed-effects logistic models as multiple regressions and using the pwr package (Champely et al., 2017) in R ($\alpha = .05$, power = 0.80), the minimum detectable R^2 values were approximately .29 for RQ1, .35 for RQ2, and .42 for RQ3. Full details of the analyses are available on the Open Science Framework (https://osf.io/zfw9d/overview?view_only=297a89b3637a4431bec57534a653fe84).

Results

Descriptive statistics

We report descriptive statistics to provide an overview of nonverbal communication patterns in mother-child interactions across ages and genders. Table 2 presents the prevalence of cues by age, gender, and initiator (child or mother). Look Cues were most prevalent across all groups (46.55% to 78.8%), followed by Tactile Cues involving objects.

Table 2.

Percentage of cues preceding conversational block (with mean and standard deviation) across cue types, age, initiator, child's sex in mother-child dyadic interaction.

		Cue Percentage, <i>M</i> (<i>SD</i>)					Child's Sex (<i>N</i> _{cues})
Age (mean month)	Cue Initiator	Gestural	Look	Pointing	Tactile (Human)	Tactile (Object)	

8	Child	3.03 (17.41)	78.79 (41.51)	0.00 (0.00)	15.15 (36.41)	75.76 (43.52)	Girls (N = 33)
		4.55 (20.99)	60.61 (49.24)	0.00 (0.00)	21.21 (41.19)	51.52 (50.36)	Boys (N = 66)
	Mother	3.66 (18.81)	69.11 (46.30)	0.00 (0.00)	13.01 (33.71)	63.41 (48.27)	Girls (N = 246)
		2.85 (16.66)	69.08 (46.27)	0.44 (6.62)	21.71 (41.27)	61.62 (48.68)	Boys (N = 456)
14	Child	2.00 (14.14)	72.00 (45.36)	0.00 (0.00)	12.00 (32.83)	66.00 (47.85)	Girls (N = 50)
		2.78 (16.55)	52.78 (50.27)	0.00 (0.00)	4.17 (20.12)	50.00 (50.35)	Boys (N = 72)
	Mother	0.38 (6.14)	61.89 (48.66)	0.38 (6.14)	5.28 (22.41)	59.25 (49.23)	Girls (N = 265)
		2.07 (14.26)	62.07 (48.61)	0.00 (0.00)	10.34 (30.51)	58.97 (49.27)	Boys (N = 290)
27	Child	1.72 (13.13)	46.55 (50.32)	0.00 (0.00)	5.17 (22.34)	43.10 (49.95)	Girls (N = 58)
		0.00 (0.00)	64.84 (47.93)	0.78 (8.84)	5.47 (22.83)	59.38 (49.31)	Boys (N = 128)
	Mother	2.20 (14.74)	48.35 (50.25)	0.00 (0.00)	1.10 (10.48)	47.25 (50.20)	Girls (N = 91)
		1.06 (10.24)	59.15 (49.24)	1.06 (10.24)	5.99 (23.76)	53.87 (49.94)	Boys (N = 284)

RQ1: Prevalence of multimodal, nonverbal cues

To examine how often multimodal cues (gestural, look, pointing, tactile with human, tactile with object) precede conversational blocks, we fitted a binomial linear mixed effect model as follows, with initiator, age and child's sex as covariates:

$$\text{cue presence (0, 1)} \sim \text{multimodal cues} + \\ \text{initiator (mother, child)} + \text{age group (8, 14, 27 months)} + \text{sex (girls, boys)} + (1 \mid \text{subject})$$

The marginal R^2 of the fitted models were .64. The analysis of deviance of the model (Intercept, $\chi^2(1) = 318.67, p < .001$) showed that the presence of cues was significantly

predicted by the type of cue, $\chi^2(4) = 1885.14$, $p < .001$, and age group, $\chi^2(2) = 18.75$, $p < .001$, but not initiator, $\chi^2(1) = 0.11$, $p = .745$ and sex, $\chi^2(1) = 0.20$, $p = .651$. Pairwise comparisons of cue category showed that emmean (estimated marginal means) probability of look cues (61.62%) and tactile cues involving objects (56.43%) were the most prevalent, followed by tactile cues involving humans (9.94%), gestural cues (1.76%), and pointing cues (0.29%). Pairwise comparisons revealed significant differences between all cue types ($p < .009$). These findings suggest that multimodal cues, particularly those involving gaze and object manipulation, most likely precede conversational blocks.

Other covariates, such as the initiator of the conversation ($p = .744$), and the child's sex, ($p = .651$), did not have significant effects on cue presence. Pairwise comparisons of age groups showed that the youngest age group ($M = 8$ months) had a significantly higher probability of cue presence (12.47%) compared to both the middle ($M = 14$ months; 9.23%, $p = .049$) and oldest age groups ($M = 27$ months; 7.17%, $p < .001$). The odds of cue presence were 1.40 times higher in the youngest group compared to the middle group and 1.85 times higher compared to the oldest group. However, the difference between the middle and oldest age groups was not statistically significant ($p = .151$).

RQ2: Developmental patterns of multimodal cue usage

To examine the developmental patterns of cue usage preceding conversational blocks, we collapsed the cues³ and coded the presence of any cue as 1. Model comparison procedures revealed that the most parsimonious model included full 3-way interaction between initiator, age and child's sex. We report the following model:

³ For finer-grained developmental patterns, we examined age-related effects separately for each cue type as a supplemental analysis of RQ2. Binomial mixed-effects models showed that look cues ($\beta = -0.027$, $p = .002$), gestural cues ($\beta = -0.078$, $p = .004$), tactile human cues ($\beta = -0.090$, $p < .001$), and tactile object cues ($\beta = -0.021$, $p = .011$) all decreased significantly with age, whereas pointing cues showed no reliable age effect ($\beta = 0.013$, $p = .94$). These results confirm that the overall decline in multimodal cue usage reflects a broad reduction across multiple nonverbal cue types.

$$\begin{aligned} &\text{cue presence (0, 1)} \sim \\ &\text{initiator (mother, child)} \times \text{age (months)} \times \text{sex (girls, boys)} \\ &+ (1 \mid \text{subject}) \end{aligned}$$

It is noteworthy that the cue categories were collapsed, and cue presence was redefined as follows: whenever any cue was present, it was coded as 1, and when there were no nonverbal cues preceding the conversational blocks, it was coded as 0. The results of the model are presented in Table 3. The marginal R^2 of the fitted models were .01.

Table 3
Analysis of Deviance Table (Type III Wald Chi-Square Tests) for Developmental Patterns of Cue Presence.

Predictor	χ^2	df	p
(Intercept)	2.49	1	.115
Initiator	4.21	1	.040*
Age	14.49	1	< .001***
Sex	13.63	1	< .001***
Initiator $\times$ Age	2.97	1	.085
Initiator $\times$ Sex	10.11	1	.001**
Age $\times$ Sex	13.76	1	< .001***
Initiator $\times$ Age $\times$ Sex	7.12	1	.008**

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

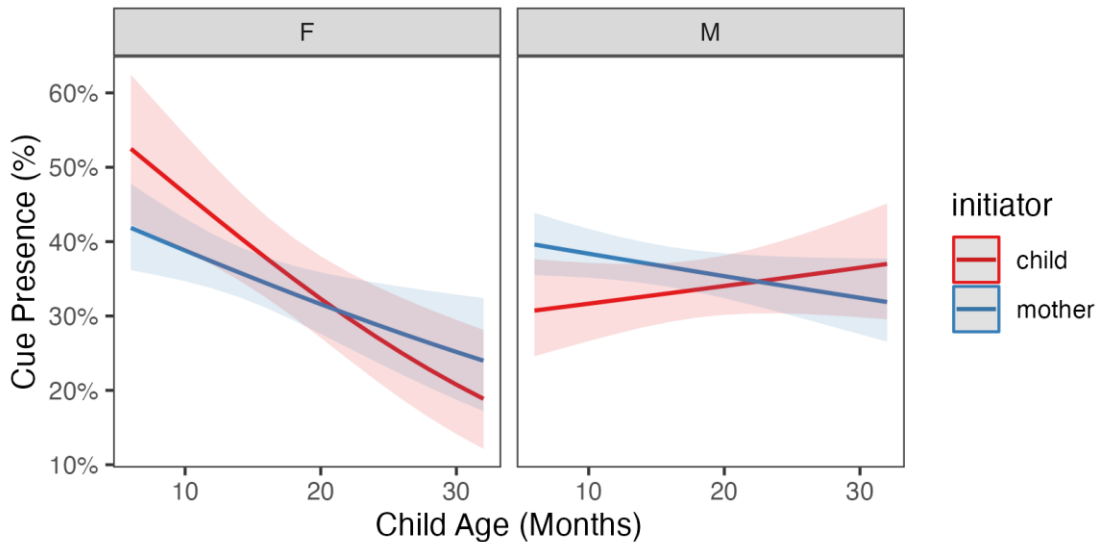
As depicted in Figure 3, we observe significant trends in cue presence across child age and sex preceding conversation initiated by mothers and children. For female children (panel F), there is a decreasing trend in cue presence as the child's age increases, in conversations initiated by the mother or the child. In the case of male children (panel M⁴), the cue presence in mother-initiated conversations shows a decreasing trend as the child's age

⁴ It remained unclear whether the increase in male children's nonverbal cues in child-initiated conversation was the result of significant increase in particular cues (e.g., action-related vs non-action-related). Further analyses suggested no evidence of cue-type-specific and cue-actor-specific changes with age.

increases, while the cue presence in child-initiated conversations increases with age. These findings suggest that their communication strategies adapt based on the child's age and sex.

Figure 3.

Percentage of cue presence preceding mother-child conversation by child's age, sex (F = girls; M = boys), and initiator.



RQ3: Role of initiator in cue usage

We investigate whether the initiator of a conversation is also more likely to produce multimodal cues preceding their own initiated conversation. Model comparison procedures revealed that the most parsimonious model included interactions between initiator type (mother, or child), cue actor (mother, child, or mother & child) and ages. We report the following model⁵:

$$\begin{aligned} &\text{cue presence (0, 1)} \sim \\ &\quad \text{initiator} \times \text{cue actor} \times \text{age} + \\ &\quad \text{initiator} \times \text{age} \times \text{sex} + (1 \mid \text{subject}) \end{aligned}$$

⁵ Adding sex as an additional interaction term to Initiator $\times$ Cue Actor $\times$ Age did not significantly improve the model.

Note that the cue actor variable was coded with the subcategory of the main cue types by filtering cue labels that contained cue actor information. The results of the model are presented in Table 4. The marginal R^2 of the fitted models were .04.

Table 4.
Analysis of Deviance Table (Type III Wald Chi-Square Tests) for Role of Initiator in Cue Usage.

Predictor	χ^2	<i>df</i>	<i>p</i>
(Intercept)	3.29	1	.070
Initiator	9.14	1	.002**
Cue Actor	1.34	2	.513
Age	5.26	1	.022*
Sex	14.39	1	< .001***
Initiator × Cue Actor	4.43	2	.109
Initiator × Age	0.71	1	.400
Cue Actor × Age	6.42	2	.040*
Initiator × Sex	10.87	1	.001**
Age × Sex	14.62	1	< .001***
Initiator × Cue Actor × Age	8.22	2	.016*
Initiator × Age × Sex ^a	7.93	1	.005**

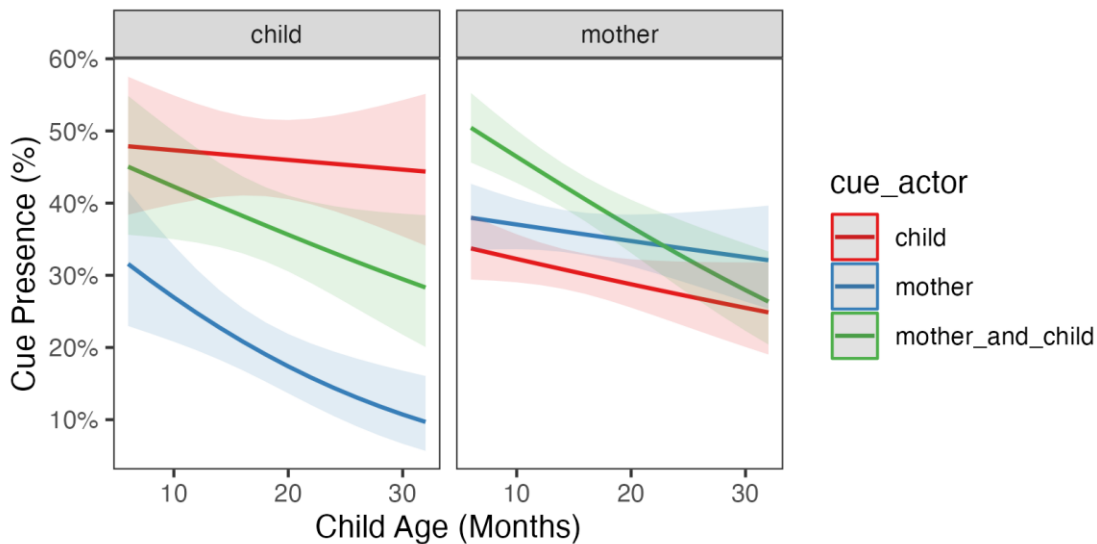
^a It is noteworthy that the interaction between initiator, age and sex is robust and the trend observed remained consistent with Model 1.

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

The pairwise comparisons between initiator, cue actor and age revealed significant differences in cue presence based on the initiator (child or mother), the cue actor (child, mother, or mother and child) and developmental pattern. When the initiator was the child, the probability of cue presence was significantly higher for the child (*probability* = 46.6%, *SE* = 2.90) compared to the mother (*probability* = 21.7%, *SE* = 2.29, $p < .001$), but there were no differences with mother and child as joint cue actors (*probability* = 38.8%, *SE* = 2.82, $p = .097$). When the initiator was the mother, the probability of cue presence was significantly higher for the mother condition (*probability* = 35.9%, *SE* = 1.48) compared to the child (*probability* = 30.4%, *SE* = 1.40, $p = .003$), but lower than the mother and child (*probability* =

41.4%, $SE = 1.55$, $p = .004$) condition. The graph depicted in Figure 4 shows that the pattern of cue presence changes depending on who initiates the conversation.

Figure 4.
Developmental pattern of cue presence changes among cue actors, with separate panels for each conversational block initiator.



For child-initiated exchanges, the cues produced by the child maintain a relatively stable and high probability of cue presence across the age range, indicating a consistent use of multimodal cues, in contrast with cues produced by the mother, and mother and child, which showed a steeper decline with age. For mother-initiated exchanges, the cues produced by child and mother showed a decrease in cue presence with increasing child age, with the cues jointly produced by mother and child demonstrating the steepest decline.

These nuanced results of cue presence across different initiators, actors and developmental stages provide valuable insights into the interplay between verbal and nonverbal communication in early childhood. The results suggest that children and mothers are more likely to use multimodal cues when they initiate conversational exchanges, and that

mothers further adapt their cue use according to the child's developmental stage, beyond differences attributable to initiator role⁶.

Discussion

Developmental dynamics of multimodal conversational initiation

This study provides new insights into how multimodal communication supports early language development. One key result is that nonverbal cue use declines as children grow older, consistent with the idea that improving verbal skills reduces the need for nonverbal scaffolding (Ko et al., 2023; West & Iverson, 2021). This developmental shift aligns with previous research highlighting the importance of joint attentional processes (Tomasello, 1988) and the multimodality of early linguistic environments (Custode & Tamis-LeMonda, 2020; Gogate et al., 2000; Tamis-LeMonda et al., 2013) in facilitating language acquisition and social understanding. The pronounced reduction in initiatory nonverbal cues within mother-initiated conversational blocks suggests that as children become more proficient in language, mothers may adapt their communication strategies, reducing their use of nonverbal cues and modeling more verbal communication (Ko et al., 2023; Vigliocco et al., 2014).

Another notable finding is the initiator effect: both mothers and children used more multimodal cues immediately before conversational blocks they themselves initiated. This pattern suggests that participants deploy gaze, touch, and gesture strategically to secure a partner's attention and establish the frame for interaction. These results are consistent with prior research showing increased engagement and responsiveness in self-initiated exchanges (Ko et al., 2016; Chai et al., 2025) and highlight the adaptive nature of these strategies for

⁶ A post-hoc sensitivity analysis indicated that, given the current sample size ($\alpha = .05$, power = 0.80), the models were only sensitive to relatively large fixed-effect patterns, with minimum detectable R^2 of approximately .29 (RQ1), .35 (RQ2), and .42 (RQ3). In comparison, the marginal R^2 of the fitted models were .64 (theoretical) / .53 (delta) for RQ1, .01 / .008 for RQ2, and .04 / .03 for RQ3, suggesting that while the large cue-type and age-group effects in RQ1 were well within the detectable range, the much smaller developmental and initiator-cue-actor patterns in RQ2 and RQ3 fall below the sensitivity of the design and may therefore be underpowered.

launching conversations. Such findings extend earlier work on the role of gaze and gesture in conversational entry (Cary, 1978; Ho et al., 2015; Kendon, 1967; Duncan Jr., 1969; Wiener et al., 1972) by showing that strategic cue use is not limited to adults but also evident in young children.

Cue-type distributions and material engagement in early interaction

Beyond overall frequency and timing, the distribution of specific cue types provides further insight into how early communication is organized. Look cues and object-related tactile cues were the most frequent signals observed before conversational blocks, underscoring the importance of joint attention and material engagement in interaction. This finding is consistent with previous research linking the role of eye gaze with joint attention (Çetinçelik et al., 2021; Csibra, 2010; Gredebäck et al., 2010; Yu & Ballard, 2007) and the use of object manipulation to language learning (Abu-Zhaya et al., 2017; Ko et al., 2023). From a 5E cognition perspective (Newen et al., 2018), these results illustrate how communication is embodied (e.g., gaze direction, hand movements), embedded in material and social contexts, and extended through the use of physical objects. Rather than serving merely as supplements to speech, these cues appear to form the foundational scaffolding on which symbolic language later develops. The prevalence of object manipulation as an "anchor" for conversation particularly resonates with Material Engagement Theory (Malafouris, 2013), which posits that cognitive and social processes are materially grounded, with objects serving as active participants in meaning-making.

Operationalization of gesture and tactile-object cues

Our finding that gestural cues were relatively infrequent compared to look and tactile-object cues warrants careful interpretation in light of how gestures were operationalized in this study. In this study, gestural cues were defined as visible movements directed toward objects or actions within proximal reach, explicitly excluding sustained touching or holding,

which were classified as tactile cues. This classification meant that ostensive actions, showing and giving, were coded as gestures only up to the moment sustained contact occurred; once the infant or mother held, maintained contact with, or offered the object to their partner, the behavior transitioned to our tactile-object cue category. Recent developmental research emphasizes that ostensive gestures (showing, giving, offering) are among the earliest communicative gestures to emerge, often preceding pointing, and play critical roles in establishing joint attention, material engagement, and triadic interaction during the first year and beyond (Guevara et al., 2024; Cameron-Faulkner et al., 2015; Moreno-Núñez et al., 2020; Choi et al., 2021). The high prevalence of tactile-object cues in our data may therefore partly reflect the contact and sustained-manipulation phases of ostensive actions, suggesting that object-mediated communication, integrating both the gestural (attention-directing) and tactile (material-engagement) phases, forms a cohesive system for initiating and scaffolding early conversational exchanges. To more fully capture the developmental patterns and communicative functions of ostensive gestures, future research should code showing, giving, and offering as a distinct category, potentially differentiating their pre-contact and contact phases.

Multimodal initiation as embodied, enacted, and extended cognition

Despite these limitations, this study contributes to the 5E cognition framework (embodied, embedded, enacted, extended, ecological; Newen et al., 2018) by demonstrating how multimodal communication in mother-child dyads exemplifies materially grounded, socially distributed cognition. The prominence of look cues (61.62%) and object-related tactile cues (56.43%) in conversation initiation reveals how meaning emerges through embodied coordination of gaze, touch, and vocalization rather than through isolated cognitive processes. The observed age-related decline in nonverbal cue usage: from 12.47% in infants under 12 months of age to 7.17% by 24 months of age, illustrates the enacted nature of

communicative development, where verbal competence reorganizes rather than replaces embodied interaction strategies. Crucially, the initiator effect, whereby both mothers and children deploy more multimodal cues in self-initiated exchanges, demonstrates extended cognition through strategic coordination of bodies, objects, and shared attention. From a Material Engagement Theory perspective (Malafouris, 2013), these findings show how objects function as cognitive resources that expand communicative potential, with the ecological situatedness of these interactions within culturally structured environments scaffolding increasingly sophisticated forms of joint attention and symbolic understanding.

Individual variation and implications for communicative development

The persistence of multimodal cues even as children develop linguistic competence raises important questions about individual variation in communicative development. Although linguistic ability was not a primary focus of the present study, we conducted supplementary analyses examining whether children's receptive and expressive vocabulary ability contributes to multimodal cue use beyond chronological age. These analyses indicated that apparent negative associations between some tactile cues and vocabulary were largely driven by developmental factors: younger children both relied more heavily on human-directed touch and had smaller vocabularies. Once age was statistically controlled, tactile cue use was no longer independently associated with vocabulary outcomes (cf. Author, 2025, which shows that touch–speech alignment, rather than touch frequency, predicts receptive vocabulary in preverbal infants). Full details of these supplementary analyses are available on the Open Science Framework (https://osf.io/zfw9d/overview?view_only=297a89b3637a4431bec57534a653fe84).

At the same time, the high variability observed during this developmental period suggests that individual differences in vocabulary size, utterance length, or verbal productivity may still shape how children deploy multimodal resources to initiate interaction.

Such differences may interact with gendered parental expectations and culturally specific interaction styles: for example, if boys receive less verbal scaffolding and more physical or object-oriented engagement (as noted in the Korean context), they may maintain higher reliance on manual and tactile cues. Within a 5E framework, these patterns highlight how linguistic development, material engagement, cultural practices, and individual variation are mutually constitutive: children's emerging verbal abilities, caregiver expectations, and embodied affordances jointly shape the local ecology of communicative initiation. Future research integrating standardized language measures or detailed lexical counts with multimodal coding would be valuable for testing these hypotheses.

Sex-differentiated multimodal development: exploratory evidence

Child sex emerged as a significant moderator only in post hoc model comparisons and was not part of the original analytic plan; accordingly, these patterns should be more appropriately interpreted as exploratory rather than as evidence of systematic developmental differences. In the present sample, girls showed a gradual decline in multimodal cue use across age for both mother- and child-initiated blocks, whereas boys showed a decline in mother-initiated blocks but a relative increase in cue presence in child-initiated blocks with age. Given the modest sample size and the fact that the relevant mixed-effects models explained only a small proportion of variance, we regard these sex-related trends as hypothesis-generating signals that require independent confirmation in larger, pre-registered datasets.

Within this cautious framing, one possibility is that the observed trajectories reflect differences in how boys and girls coordinate multimodal resources as they take the lead in interaction, resonating with prior reports of sex-related variation in early language development (Eriksson et al., 2012; Barel & Tzischinsky, 2018). Alternatively, children's opportunities and strategies for multimodal communication over time might be shaped by

parents' tendency to engage more physically with boys and more verbally with girls (Endendijk et al., 2016; Mascaro et al., 2017). These interpretations should also be considered within the sociocultural context of contemporary Korean parenting, where overall levels of engagement and responsiveness appear broadly comparable to Western samples yet some gendered expectations persist; against this mixed backdrop, it remains difficult to determine whether the apparent sex-related differences observed here reflect stable developmental tendencies, sociocultural variation, or sample-specific effects, and we therefore treat them as context-bound and in need of replication before substantive conclusions are drawn.

Methodological considerations and ecological constraints

Several methodological considerations temper the interpretation of these findings. A strength of this study is its naturalistic free-play design in which interactional trajectories emerge through mothers' and children' own activities (Ko et al., 2020; Tamis-LeMonda et al., 2017; Wojcik et al., 2024). In this methodological tradition, "naturalistic" refers to interactional openness rather than to the physical setting alone, such that mothers and children can initiate interactions, coordinate attention, and deploy nonverbal resources without experimenter-imposed tasks or scripts. This approach is well suited to developing culturally sensitive descriptions of early communicative repertoires because it captures how interaction is organised under participant-driven conditions rather than laboratory-imposed assessment constraints (Kirova & Hennig, 2013; Rao et al., 2020).

At the same time, several constraints should be considered when interpreting these findings. The sample comprised 35 Korean mother–child dyads drawn from a relatively homogeneous group of highly educated mothers, which limits the extent to which the findings can be extended to more diverse socioeconomic and cultural contexts. The cross-sectional design, while efficient for identifying age-related patterns, offers discrete

developmental snapshots and does not capture transitional periods or individual developmental trajectories. Within this design, individual differences and contextual influences that unfold over longer timescales are likely to be underrepresented, contributing to substantial unexplained variance in the models. This is particularly relevant for the exploratory sex-difference analyses, where the modest sample size and cross-sectional structure further constrain the generalisability of the estimated effects.

Methodologically, the naturalistic approach involved trade-offs between interactional openness and experimental control, as the semi-controlled laboratory setting may have influenced behavior differently than home environments. Because conversational initiation is sensitive to the number, age, and identity of interaction partners, these design features substantially constrain the interpretive scope of our findings with respect to other caregivers (e.g., fathers), siblings, peers, and broader family contexts. Accordingly, the present results should be interpreted as characterizing multimodal initiation patterns in mother–infant free-play under controlled naturalistic conditions, offering a theoretically informative case of how embodied resources are organized in early interaction rather than as representative of everyday family interactions across a broader range of partners and settings.

The coding protocol, while achieving satisfactory inter-rater reliability ($\kappa = 0.806$), focused on predetermined cue categories within 3-second temporal windows, potentially overlooking other embodied or materially mediated behaviors occurring over longer interactional timescales. The exclusive focus on mother–child dyads and the corpus-based design, while enabling fine-grained behavioral analysis, limits certain statistical analyses and controlled variable manipulations that could strengthen causal inference.

Conclusion

This study shows that conversational initiation between mothers and children is supported by multimodal cues, particularly gaze and object-related touch, which serve to secure

attention and establish shared focus before speech begins. The frequency of these cues decreases as children mature and rely more on verbal strategies, revealing a developmental shift from sensorimotor to symbolic means of initiating interaction. Both mothers and children used more multimodal cues when they initiated conversations, suggesting that these signals are deployed strategically to manage interactional entry. These findings extend the 5E cognition framework by showing how early conversational exchanges are embodied, embedded, enacted, extended, and ecologically situated, with material engagement, especially through object manipulation, as a key resource for communication. They also show that caregiver–child strategies adapt across developmental stages and that early communication is not simply a precursor to language, but a dynamic, distributed process that reorganizes as children gain verbal competence. This developmental perspective provides insights into how multimodal scaffolding supports early language development and offers a foundation for designing interventions that promote effective caregiver–child communication.

References

- Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, 44(5), 1088–1116. <https://doi.org/10.1017/S0305000916000416>
- Barel, E., & Tzischinsky, O. (2018). Age and Sex Differences in Verbal and Visuospatial Abilities. *Advances in Cognitive Psychology*, 2(14), 51–61. <https://doi.org/10.5709/acp-0238-x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, 30(3), 359–375. <https://doi.org/10.1111/j.2044-835X.2011.02043.x>
- Bornstein, M. H., Cote, L. R., & Kwak, K. (2019). Comparative and Individual Perspectives on Mother–Infant Interactions with People and Objects among South Koreans, Korean Americans, and European Americans. *Infancy*, 24(4), 526–546. <https://doi.org/10.1111/inf.12288>
- Bruner, J. (1985). Child's Talk: Learning to Use Language. *Child Language Teaching and Therapy*, 1(1), 111–114. <https://doi.org/10.1177/026565908500100113>
- Cary, M. S. (1978). The Role of Gaze in the Initiation of Conversation. *Social Psychology*, 41(3), 269–271. <https://doi.org/10.2307/3033565>
- Caselli, M. C., Rinaldi, P., Stefanini, S., & Volterra, V. (2012). Early Action and Gesture “Vocabulary” and Its Relation With Word Comprehension and Production. *Child Development*, 83(2), 526–542. <https://doi.org/10.1111/j.1467-8624.2011.01727.x>

- Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.589096>
- Chai, J. H., Jung, J., & Ko, E.-S. (2025). The dynamics of initiation in caregiver–child conversational interactions. *Journal of Child Language*, 52(4), 839–860. <https://doi.org/10.1017/S030500092400062X>
- [Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., ... De Rosario, H. \(2017\). pwr: Basic functions for power analysis \(R package\).](#)
- Chong, S. C. F., Werker, J. F., Russell, J.A. & Carroll, J. M. (2003), Three facial expressions mothers direct to their infants. *Infant and Child Development*, 12: 211-232. <https://doi.org/10.1002/icd.286>
- Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, 25(2), 141–168. <https://doi.org/10.1111/j.1468-0017.2009.01384.x>
- Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, 25(6), 809–826. <https://doi.org/10.1111/inf.12361>
- Egeren, L., Barratt, M., & Roach, M. (2001). Mother–Infant Responsiveness: Timing, Mutual Regulation, and Interactional Context. *Developmental Psychology*, 37, 684–697. <https://doi.org/10.1037/0012-1649.37.5.684>
- Endendijk, J. J., Groeneveld, M. G., Bakermans-Kranenburg, M. J., & Mesman, J. (2016). Gender-differentiated parenting revisited: Meta-analysis reveals very few differences in parental control of boys and girls. *PloS one*, 11(7), e0159193.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., ... & Gallego, C. (2012). Differences between girls and boys in emerging language

- skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30(2), 326-343.
- Duncan Jr., S. (1969). Nonverbal communication. *Psychological Bulletin*, 72(2), 118–137.
<https://doi.org/10.1037/h0027795>
- Faul, F., Erdfelder, E., Lang, A.G. et al. G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences, *Behavior Research Methods* 39, 175–191 (2007). <https://doi.org/10.3758/BF03193146>
- Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.
- Gallagher, S. (2023). Embodied and Enactive Approaches to Cognition. *Elements in Philosophy of Mind*. <https://doi.org/10.1017/9781009209793>
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A Study of Multimodal Motherese: The Role of Temporal Synchrony between Verbal Labels and Gestures. *Child Development*, 71(4), 878–894. <https://doi.org/10.1111/1467-8624.00197>
- Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. M. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, 10(6), 778–785.
- Goldin-Meadow, S. (2009). How gesture promotes learning. *Trends in Cognitive Sciences*, 13(10), 448–454.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489–1522.
- Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse & Society*, 18(1), 53–73.
- Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental Neuropsychology*, 35(1), 1–19.
<https://doi.org/10.1080/87565640903325758>

- Han, N. (1999). Understanding of contemporary Korean family [현대 한국 가족의 이해]. Seoul, South Korea: Ilji-sa.
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PLOS ONE*, 10(8), e0136905. <https://doi.org/10.1371/journal.pone.0136905>
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–371.
- Jeong, Y., & Kwak, K. (2005). Development of Coordinated joint attention in infancy: looking through attentional state and pointing behavior in mother-infant interaction. *The Korean Journal of Developmental Psychology*, 18(1), 137-154.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kirova, A., & Hennig, K. (2013). Culturally Responsive Assessment Practices: Examples from an Intercultural Multilingual Early Learning Program for Newcomer Children. *Power and Education*, 5(2), 106–119. <https://doi.org/10.2304/power.2013.5.2.106>
- Kosie, J. E., & Lew-Williams, C. (2024). Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Developmental Science*, 27, e13515. <https://doi.org/10.1111/desc.13515>
- Ko, E.-S., Abu-Zhaya, R., Kim, E.-S., Kim, T., On, K.-W., Kim, H., Zhang, B.-T., & Seidl, A. (2023). Mothers' use of touch across infants' development and its implications for word learning: Evidence from Korean dyadic interactions. *Infancy*, 28(3), 597–618. <https://doi.org/10.1111/infa.12532>
- Ko, E. S., Jo, J., On, K. W., & Zhang, B. T. (2020). Introducing the Ko Corpus of Korean mother–child interaction. *Frontiers in psychology*, 11, 602623.

- Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, 43(2), 284–309. <https://doi.org/10.1017/S0305000915000203>
- Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., Blair, C. B., & Family Life Project Key Investigators (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development*, 85(5), 1898–1914. <https://doi.org/10.1111/cdev.12249>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://cran.r-project.org/package=emmeans>
- Lüdtke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26), 772. <https://doi.org/10.21105/joss.00772>
- Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Patil, I., Thériault, R., & Lüdtke, D. (2025). modelbased: An R package to make the most out of your statistical models through marginal means, marginal effects, and model predictions. *Journal of Open Source Software*, 10(109), 7969. <https://doi.org/10.21105/joss.07969>
- Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. The MIT Press. <https://doi.org/10.7551/mitpress/9476.001.0001>
- Malafouris, L. (2020). Mind and material engagement. *Phenomenology and the Cognitive Sciences*, 19(1), 1–17.
- Malafouris, L. (2020). Thinking as “Thinging”: Psychology With Things. *Current Directions in Psychological Science*, 29(1), 3–8. <https://doi.org/10.1177/0963721419873349>

- Mascaro, J. S., Rentscher, K. E., Hackett, P. D., Mehl, M. R., & Rilling, J. K. (2017). Child gender influences paternal behavior, language, and brain function. *Behavioral Neuroscience*, 131(3), 262.
- Moore, C., & D'Entremont, B. (2001). Developmental Changes in Pointing as a Function of Attentional Focus. *Journal of Cognition and Development*, 2(2), 109–129.
https://doi.org/10.1207/S15327647JCD0202_1
- Moreno-Núñez, A., Murillo, E., Casla, M., & Rujas, I. (2021). The multimodality of infant's rhythmic movements as a modulator of the interaction with their caregivers. *Infant Behavior and Development*, 65, 101645.
<https://doi.org/10.1016/j.infbeh.2021.101645>
- Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, 16(5), 269–274. <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
- Newen, A., Bruin, L. D., & Gallagher, S. (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press.
- Ochs, E. (1982). Talking to children in Western Samoa. *Language in Society*, 11(1), 77–104.
- Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge University Press.
- Park, S.-Y., & Cheah, C. S. L. (2005). Korean mothers' proactive socialisation beliefs regarding preschoolers' social skills. *International Journal of Behavioral Development*, 29(1), 24–34. <https://doi.org/10.1080/01650250444000306>
- Park, S.-Y., & Kim, M.-J. (2006). Korean Mothers' Ideal and Actual Parenting Behaviors Toward their Young Children as a Function of Child Gender, Age, and Birth Order. *International Journal of Human Ecology*, 7.

- Park, S. Y., Lee, S., & Park, U. (2000). The relations between mothers' value and parenting attitudes and children's behavioural inhibition. *Journal of Korean Home Management Association*, 18, 85–95.
- Park, S. Y., Soe, S. J., & Bornstein, M. H. (2005). Mother-Infant Interaction Styles Associated with Infant Development. *Korean Journal of Child Studies*, 26(5), 15–30.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rao, N., Mirpuri, S., Sincovich, A., & Brinkman, S. A. (2020). Overcoming challenges in measuring early childhood development across cultures. *The Lancet Child & Adolescent Health*, 4(5), 352–354. [https://doi.org/10.1016/S2352-4642\(20\)30026-2](https://doi.org/10.1016/S2352-4642(20)30026-2)
- Rietveld, E., Denys, D., & Van Westen, M. (2018). Ecological-Enactive Cognition as engaging with a field of relevant affordances: The Skilled Intentionality Framework (SIF). In A. Newen, L. De Bruin, & S. Gallagher (Eds), *The Oxford Handbook of 4E Cognition* (1st edn, pp. 41–70). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198735410.013.3>
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.
- Salo, V. C., King, L. S., Gotlib, I. H., & Humphreys, K. L. (2022). Infants who experience more adult-initiated conversations have better expressive language in toddlerhood. *Infancy*, 27(5), 916–936. <https://doi.org/10.1111/infa.12487>
- Schachner, A., Hannon, E. E., & Bar, M. (2022). Prosodic faces: Caregivers' facial expressions during infant-directed speech. *Scientific Reports*, 12, 22311.
<https://doi.org/10.1038/s41598-022-26981-7>
- Schieffelin, B. B. (1990). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge University Press.

- Shin, S. Y., Rowe, M. L., & Lee, H. S. (2024). Early gesture use predicts children's language development in South Korea: New evidence supporting the cross-cultural importance of pointing. *Infancy*, 29(3), 327–354. <https://doi.org/10.1111/inf.12585>
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4(1), 1–22.
- Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, 20(6), e12456. <https://doi.org/10.1111/desc.12456>
- Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From Action to Interaction: Infant Object Exploration and Mothers' Contingent Responsiveness. *IEEE Transactions on Autonomous Mental Development*, 5(3), 202–209. <https://doi.org/10.1109/TAMD.2013.2269905>
- Tamis-LeMonda, C. S., & Masek, L. R. (2023). Embodied and embedded learning: Child, caregiver, and context. *Current Directions in Psychological Science*, 32(5), 369–378. <https://doi.org/10.1177/0963721423117873>
- Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, 15(2), 103–125. <https://doi.org/10.1080/15475441.2018.1533472>
- Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences*, 10(1), 69–88. [https://doi.org/10.1016/0388-0001\(88\)90006-X](https://doi.org/10.1016/0388-0001(88)90006-X)
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691. <https://doi.org/10.1017/S0140525X05000129>

- Vesper, C., & Sevdalis, V. (2020). Informing, Coordinating, and Performing: A Perspective on Functions of Sensorimotor Communication. *Frontiers in Human Neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.00168>
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130292. <https://doi.org/10.1098/rstb.2013.0292>
- Vygotsky, L. S., & Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.
- West, K. L., & Iverson, J. M. (2021). Communication changes when infants begin to walk. *Developmental Science*, 24(5), e13102. <https://doi.org/10.1111/desc.13102>
- Wiener, M., Devoe, S., Rubinow, S., & Geller, J. (1972). Nonverbal behavior and nonverbal communication. *Psychological Review*, 79(3), 185–214. <https://doi.org/10.1037/h0032710>
- Wojcik, E. H., Pierce, M. C., Stevens, G., & Goulding, S. J. (2024). Referent-oriented interactions in infancy: A naturalistic, longitudinal case study from an English-speaking household. *Infant Behavior and Development*, 74, 101911. <https://doi.org/10.1016/j.infbeh.2023.101911>
- Yoon, H., & Choi, Y. (2023). Cultural Influence on Parenting Behavior: Distal Parenting Strategies Are Dominant in Korean Mother-Infant Interactions. *Korean Journal of Developmental Psychology*, 36(4), 83–101. <https://doi.org/10.35574/KJDP.2023.12.36.4.83>
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70(13), 2149–2165. <https://doi.org/10.1016/j.neucom.2006.01.034>

Yurkovic, J. R., Kennedy, D. P., & Yu, C. (2021). Multimodal Behaviors from Children

Elicit Parent Responses in Real-Time Social Interaction. *Proceedings of the Annual*

Meeting of the Cognitive Science Society, 43(43).

<https://escholarship.org/uc/item/2qw0d216>